Review of sampling hard-to-reach and hidden populations for HIV surveillance

Robert Magnani[a], Keith Sabin[b], Tobi Saidel[a] and Douglas Heckathorn[c]

Adequate surveillance of hard-to-reach and 'hidden' subpopulations is crucial to containing the HIV epidemic in low prevalence settings and in slowing the rate of transmission in high prevalence settings. For a variety of reasons, however, conventional facility and survey-based surveillance data collection strategies are ineffective for a number of key subpopulations, particularly those whose behaviors are illegal or illicit. This paper critically reviews alternative sampling strategies for undertaking behavioral or biological surveillance surveys of such groups. Non-probability sampling approaches such as facility-based sentinel surveillance and snowball sampling are the simplest to carry out, but are subject to a high risk of sampling/selection bias. Most of the probability sampling methods considered are limited in that they are adequate only under certain circumstances and for some groups. One relatively new method, respondent-driven sampling, an adaptation of chain-referral sampling, appears to be the most promising for general applications. However, as its applicability to HIV surveillance in resource-poor settings has yet to be established, further field trials are needed before a firm conclusion can be reached.

© 2005 Lippincott Williams & Wilkins

Introduction

The main function of HIV/AIDS surveillance is to provide an understanding of local epidemics, including the source of new infections over time and the behavioral and biological factors driving infection spread in order to provide a basis for designing and evaluating appropriate interventions [1]. In order to be effective, it is crucial that surveillance efforts focus on the segments of national or community populations that play an important role in HIV transmission. Epidemiological considerations, of course, play a central role in the selection of populations for surveillance. Population subgroups should be chosen on the basis of their ability to provide information about where and among whom HIV is spreading and where and among whom the behaviors that expose individuals to HIV are being practised [2,3]. Surveillance should also focus on subpopulations that are large enough to influence the spread of HIV meaningfully.

In low–grade and concentrated epidemic settings, these groups will generally consist of sex workers, injection drug users (IDU), men who have sex with men (MSM), and specific mobile or migrating population groups. However, there is also an important rationale for conducting surveillance among these 'high–risk' groups in high HIV prevalence settings, especially when targeted interventions are planned.

A primary challenge for surveillance of these high–risk subpopulations is obtaining 'representative' samples of them for surveillance measurement purposes [4,5]. The challenge arises from the fact that many such groups are 'hidden'; that is, no sampling frame exists for them, and because the behaviors in which they engage are either illegal or illicit, they generally prefer not to participate in surveillance data collection activities [6,7]. Meaningful surveillance thus requires that sampling strategies that are both feasible and capable of producing unbiased estimates (or more realistically estimates with minimal levels of bias) be devised for population subgroups that are not efficiently 'captured' using conventional surveillance data collection strategies.

From the [a]Family Health International, Arlington, VA, USA, the [b]Centers for Disease Control and Prevention, Atlanta, GA, USA, and the [c]Cornell University, Ithaca, NY, USA.

Correspondence to Keith Sabin, MS E-30, 1600 Clifton Road NE, Global AIDS Program, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.

Tel: +1 404 639 6314; e-mail: ksabin@cdc.gov

This paper critically reviews the sampling approaches that have and potentially could be used in 'second-generation' HIV surveillance efforts for such groups. The paper begins with a brief review of public health surveillance, which defines the context for evaluating alternative sampling approaches. We then enumerate the major challenges encountered in sampling the types of hard-to-reach and 'hidden' populations that are of interest for HIV surveillance and the merits of different sampling approaches in meeting these challenges. The paper concludes with some recommendations for further testing and research to advance the state of the science of HIV surveillance for hidden populations.

Public health surveillance

A crucial feature of public health surveillance is that the data should be representative of and thus generalizable to the population under surveillance, i.e. have high external validity. When a disease or other outcome of interest is highly prevalent in the general population and a large proportion of the population comes into at least periodic contact with health services, routine reporting by health clinics and other service-providing institutions (health clinics, hospitals, private doctor's offices, drug treatment centers, correctional facilities, family planning centers, etc.) suffices as a surveillance mechanism. Because public health resources are scarce, surveillance has traditionally targeted the 'low hanging fruit' (i.e. the most easily accessed populations) for HIV surveillance. In most developing countries, the only sources of routinely available HIV surveillance data have been pregnant women seeking antenatal care, sexually transmitted disease (STD) patients, and military recruits (the data for which are often not publicly available).

In the absence of reporting requirements or facility-based data collection capabilities, periodic general population surveys can provide adequate HIV surveillance data, albeit at a higher cost. Multistage cluster sampling procedures for general population surveys are well established and accepted [8], and survey-based public health surveillance of common health-related behaviors has been common practice for decades. With regard to HIV/AIDS, population-based prevention indicator surveys provided much of the general population behavioral surveillance data available in developing country settings in the 1990s, and both behavioral and biological HIV surveillance data are currently being obtained on a regular basis via Demographic and Health Surveys and AIDS Indicator Surveys.

However, when the risk-taking behaviors that justify the inclusion of population subgroups in HIV surveillance are stigmatized or illegal in the society at large, conventional household surveys are unlikely to produce accurate surveillance data. There are several reasons for this. First, no list or sampling frame of the subpopulations exists, and creating a useful sampling frame is generally either infeasible or prohibitively expensive. The ethnographic mapping of populations, for example, is relatively labor intensive when performed for sample frame development. Mapping also relies on readily identifiable elements of a population. One cannot map IDU who have drugs delivered and inject at home. Other applications of mapping, e.g. for targeted sampling, are essentially convenience samples. Second, because they often represent a small proportion of the general population, obtaining statistically reliable data for such subpopulations through household surveys would require prohibitively large sample sizes. Finally, because they engage in behaviors that are illegal or at least stigmatized in many settings, members of such subpopulations are often reluctant to participate in surveys and risk revealing their behaviors to others who may be present at the time of a survey.

The existence of hidden populations presents a dilemma for HIV surveillance, as their omission from surveillance systems leaves important gaps in our knowledge and understanding of the HIV/AIDS epidemics. The importance of these populations warrants special attention to the development of sampling methods that provide valid estimates of infection rates and behaviors among their members.

Sampling of hard-to-reach and hidden populations

Sampling procedures should be capable of reaching all members of the population or subpopulation under surveillance in order to produce unbiased estimates of trends in HIV infection rates and behavioral risks. If they are not, observed changes in behaviors and infection rates may be confounded by factors that are the result of differences in sampling procedures in successive rounds of data collection. As in other scientifically rigorous endeavors, the preferred approach for surveillance is probability sampling, defined as a sample in which sample elements are chosen randomly in such a way that each element has a non-zero probability of selection that may be calculated.

Some would argue that surveillance data do not have to be perfect (i.e. be unbiased) in order to be useful, and that undertaking surveillance that entails time-consuming and costly sampling frame development and related preparatory activities is wasteful of scarce public health resources. This argument would, in our view, have merit if the direction and magnitude of biases for the population subgroups of interest were known and were constant over time. However, because: (i) conventional surveillance

approaches often capture only a small fraction of the total population of some subgroups; (ii) the behaviors and HIV status of those covered and missed by conventional surveillance systems can differ quite substantially and in ways that cannot be reliably anticipated; and (iii) there is no assurance that biases in coverage will remain constant over time, there is the very real danger of surveillance data being misleading or failing to capture significant pockets of infection that can lead to a more generalized spread of HIV if not contained. In the light of this, it is our view that investment in obtaining higher quality surveillance data is justified, although admittedly there are limits to the magnitude of resources that can be spent on surveillance, and this reality must be considered when choosing among alternative sampling methods.

Snowball sampling

Over the past two to three decades, several methods for recruiting hidden populations for surveillance and other survey research purposes have emerged [9,10]. Perhaps the most commonly used method is snowball sampling [6–8,11–14]. Snowball sampling entails identifying an initial number of subgroup members from whom the desired data are gathered and who then serve as 'seeds', or study staff recruited respondents, to help identify other subgroup members (i.e. individuals who engage in the same types of behaviors) to be included in the sample. These individuals in turn are asked to provide information on other subgroup members, and the process continues until either a target sample size has been reached or the sample has become 'saturated' (i.e. new sample subgroup members fail to provide information that differs from that obtained from members interviewed previously).

Although initial seeds in snowball sampling are in theory randomly chosen, in practice this is difficult if not impossible to carry out. Therefore, as a practical matter, initial seeds in snowball sampling tend to be chosen via convenience sampling. Like other non-probability sampling methods, the major drawback of snowball sampling is sampling bias; that is, the danger that the sample ultimately obtained is not 'representative' of the larger population from which the sample was drawn. In snowball sampling, the sample composition is heavily influenced by the choice of initial seeds, and the method, in practice, also tends to be biased towards favoring more cooperative as opposed to randomly chosen subjects and those that are part of larger personal networks [6,7]. Non-probability sampling methods such as snowball sampling are useful in formative research and in problem definition, but are not suitable for producing data that can be confidently generalized to larger populations, although they are sometimes (incorrectly) used in this manner.

Facility-based sampling

Recruiting population members from a variety of facilities frequented by members is another commonly used method. Correctional facilities have been used to sample populations involved in illegal activities such as illicit drug use and commercial sex work [15–19]. Drug treatment centers are useful sources for finding IDU [20–22], and some STD clinics serve high proportions of MSM and commercial sex workers (CSW), with some dedicated exclusively to these populations [23–27]. Needle exchanges also provide access to IDU [28]. Each of these facilities has been used to recruit large numbers of hidden population members; however, they come with certain, similar biases. Correctional facility populations rely on the application of local laws, both the laws and the application of which can vary widely by jurisdiction. None of the options provide probability samples that can be considered representative of a given population. Individuals who have the wherewithal to obtain services, particularly in societies in which their behaviors are stigmatized, will be different from group members who do not seek and obtain these services. Furthermore, dedicated services such as STD clinics for CSW and MSM or needle exchanges and drug treatment for IDU are not common in many parts of the world. In addition, drug treatment centers offering opiate substitution therapy will not attract cocaine injectors, only heroin users.

Targeted sampling

Other sampling methods have been developed to try to overcome the limitations of snowball sampling for use with population subgroups such as those of interest for HIV behavioral and biological surveillance [29,30]. Targeted sampling, for example, extends the ideas of snowball sampling to include an initial ethnographic assessment aimed at identifying the various networks or subgroups that might exist in a given setting [31]. The subgroups so identified are then treated as sampling strata, and quota samples are chosen within each stratum using systematic sampling when feasible. The magnitude of sampling bias in targeted sampling depends on the thoroughness of the ethnographic assessment. As a practical matter, the time and resources available to undertake thorough ethnographic assessments limits the usefulness of the approach for surveillance [6,7].

Time-location sampling

Another approach that has seen increasing use in recent years takes advantage of the fact that some hidden populations tend to gather or congregate at certain types of locations [32,33]. For example, sex workers often

congregate at brothels, massage parlors, and street corners in 'red light' districts; MSM in bars and 'cruising areas' known to attract MSM; IDU at 'shooting galleries' and other locations known to be frequented by IDU. In time-location sampling (TLS), such sites are enumerated in a preliminary ethnographic mapping or presurveillance assessment exercise; the list of sites so developed is used as a sampling frame from which to choose a probability sample of sites, and data are gathered from either all or a sample of subgroup members found at the site during a pre-defined time interval (e.g. a randomly chosen 3-h time period on a randomly chosen day of the week). Because probabilities of selection can be calculated, TLS qualifies as a probability sampling method.

However, unless all or a very high percentage of sites where subgroup members congregate are identified so that they can be included in the sampling frame, and all or a very high percentage of subgroup members visit such sites at least periodically, TLS also suffers from potentially unacceptable levels of bias. Including all gathering sites can in theory be achieved given sufficient time and resources for sampling frame development, but here again there are practical limits as to the resources that can be committed to such activities on a regular basis. Because the locations where members of particular subgroups congregate change over time, it is necessary to repeat the sampling frame development exercise before each round of surveillance data collection. Having available the sampling frame from previous surveillance rounds reduces the cost of sampling frame development in subsequent rounds, but the costs of updating the sampling frame tend nevertheless to be non-trivial. As a result, there is a real danger of missing some sites, resulting in potential sampling bias.

Subgroup members who do not visit such sites pose a more serious problem. Here, no amount of rigor in constructing sampling frames of gathering sites will reduce sampling bias, and thus if a significant proportion of members of a given subgroup tend not to frequent such sites, TLS can be subject to serious sampling bias, to the extent that the behaviors and HIV status of subgroup members who do not visit gathering sites differ from those who do.

Another important source of bias with TLS is the nature of the recruitment sites. MSM attending bars and dance clubs may not want to participate in surveys in which they might learn their HIV status. IDU coming to buy drugs will want to leave as soon as possible. Sex workers working a street corner will not want to miss a potential client. Non-response will be linked closely with certain sites.

Respondent-driven sampling

The newest approach for sampling hidden populations is known as respondent-driven sampling (RDS) [6,7]. RDS has several features that allow it to overcome some of the limitations inherent in the other methods described above. The method is similar to snowball sampling in that it involves chain referral sampling. However, the recruitment process is implemented in a manner that allows for the calculation of selection probabilities, thus it qualifies as a probability sampling method [34,35]. In addition, the method has greater external validity because it is not limited to subgroup members who are accessible at sites, but rather extends the sample to all potential members of a subgroup selected for surveillance by accessing respondents through their social networks.

With RDS, 'seeds' are enlisted as temporary recruiters. They receive an explanation of the study and a limited number of coupons that can be used by them to recruit a peer who is eligible for the study. The 'seed' refers their peer to the study by providing them with a coupon that has a unique serial number. If their peer is eligible and enrolls in the study, the 'seed' may become eligible for a reimbursement for their recruitment effort. Furthermore, each referred respondent receives a similar number of coupons, as do their referred respondents, until the sample size is met. Because the referred respondent must present themselves at the study site, recruitment is entirely voluntary. Staff never need the names or contact information of potential participants.

Among the primary features that distinguish RDS from snowball sampling is that 'seeds' are limited in the number of respondents they can recruit by the number of coupons they receive (e.g. three to four), thereby minimizing the influence of initial seeds on the final sample composition. Limiting the number of recruits in this way encourages long recruitment chains, thereby increasing the 'reach' of the sample into more hidden pockets of the population. Other features that distinguish RDS from snowball sampling are that the relationship between recruiters and recruits is documented so that recruitment biases can be assessed and adjusted for in the analysis, and information on the personal network size of each respondent is collected to allow weighted analysis through 'post-stratification' to compensate for the oversampling of respondents with larger social networks. For example, a typical RDS respondent, John Doe, refers Jane Doe. Without knowing either individual's name, we ask Jane Doe for her relationship with the person who gave her the coupon. She replies casual sex partner and regular injection partner. Through the coupon serial number, we link her to John Doe with the relationship information. When John Doe returns for reimbursement, we ask his relationship to the person he referred and that information becomes linked. Furthermore, we asked John how many injectors he knows. If he responded '30', we know that Jane's theoretical probability of selection by John was one in 30.

When conducting RDS, data collection proceeds through successive 'waves' or recruitment cycles until the sample reaches 'equilibrium' with respect to the variables being measured. Equilibrium can be interpreted as a state in which the estimates converge around a stable sample composition that does not change during subsequent cycles of recruitment. In theory, equilibrium is reached within six recruitment waves or less regardless of who the initial seeds are. In addition to providing more externally valid probability samples, a major advantage of RDS is that it does not require an exhaustive mapping process to construct sampling frames. With RDS, the sampling frame is constructed during the sampling process, during which subgroup members recruit their peers and recruitment patterns are documented. Another theoretical advantage of RDS is that it is based on a dual incentive system, financial reward in combination with peer pressure, and this can be expected to reduce non-response bias because those who would not participate for financial reasons alone may do so as a favor to a friend.

Respondent-driven sampling checklist

The RDS sampling method includes four essential elements. If one or more of these is not present, then the sampling method is not RDS. These are: (i) documentation of who recruited whom must be tracked, generally through a coupon system; (ii) recruitment must be rationed with generally no more than three coupons allotted per 'seed', (iii) information on personal network size must be gathered and recorded; and (iv) recruiters and recruits must know one another (i.e. have a preexisting relationship).

Conventional cluster sampling

It should be noted that in limited circumstances, conventional cluster sampling may be an adequate sampling method for HIV surveillance of at-risk populations. For cluster sampling to be appropriate, it is necessary to have available or be possible to construct a relatively complete sampling frame of group members. Furthermore, it is necessary to be able to access all group members during the period of data collection. These requirements might be met in the case of readily accessible populations, for example, military personnel and miners. Other groups of potential interest (e.g. police, transportation workers), unless it is possible to make repeated 'callbacks' to obtain measurements from sampled group members not present at the time of data collection, could result in a potentially large non-response bias, rendering cluster sampling an infeasible option for hidden high-risk groups.

Recommendations for further testing and research

Studies validating the RDS method and comparing it with other probability sampling methods (e.g. TLS) are currently underway in a number of developing countries to assess its feasibility and utility as a sampling strategy for 'second generation' HIV surveillance. Some of the key assumptions and operational issues being investigated include: (i) how to track refusal rates and the potential impact of non-response bias; (ii) the assumption of random recruitment within personal networks; (iii) the speed with which equilibrium can be reached given the typical sample sizes and timeframes used for surveillance, and the unknown degree of overlap between networks; (iv) how initial seeds should be selected to maximize the ability to reach equilibrium in the shortest amount of time; (v) the question of appropriate incentives to maximize participation, and minimize the likelihood of refusal or the recruitment of strangers or ineligible respondents; (vi) the degree to which RDS is able to reach a portion of the population missed by other sampling methods; and (vii) how to manage multiple data collection sites, staffing and the verification of whether respondents meet inclusion criteria.

Discussion

Given the critical importance of understanding local HIV epidemics, our view is that high quality surveillance systems are very much needed. Appropriate sampling approaches are at the core of any high quality surveillance system, especially when the system is tracking populations that are 'hidden' or 'difficult to reach'. Most often when surveillance data are not interpretable or produce unexpected findings, inappropriate or inconsistent sampling methods are often the cause of the problem. These errors can result from many different sources including: (i) selection bias resulting from sampling only in selected facilities; (ii) a poor definition of the surveillance population in 'community-based' surveys; (iii) incomplete sampling frames; (iv) the use of venue-based sampling frames when many members of the population never frequent those sites; (v) an inability to locate or identify members of the subpopulation; and (vi) non-response and other sources of bias.

Tracking transmission dynamics among populations that play a critical role in the transmission of HIV is one of the many challenges we must confront if we want to improve our response to the epidemic. The methods discussed in this paper represent the best efforts to date to find feasible sampling approaches that will contribute to obtaining unbiased trends of HIV prevalence and HIV-related risk behaviors among hidden populations. Results from validation studies currently under way should begin to

provide evidence about how well these sampling methods are performing in developing country settings. Undoubtedly, there will be many lessons to share that will result in further advances in the state of HIV surveillance.

References

1.   Brown T. **Behavioral surveillance: current perspectives, and its role in catalyzing action.** *J Acquir Immune Defic Syndr* 2003; **32 (suppl. 1)**:S12–S17.
2.   Pisani E, Lazzari S, Walker N, Schwartlander B. **HIV surveillance: a global perspective.** *J Acquir Immune Defic Syndr* 2003; **32 (suppl. 1)**:S3–S11.
3.   UNAIDS. *Second generation surveillance for HIV*. Geneva: WHO and UNAIDS; 2002.
4.   Mills S, Saidel T, Bennett A, Rehle T, Hogle J, Brown T, *et al*. **HIV risk behavioral surveillance: a methodology for monitoring behavioral trends.** *AIDS* 1998; **12 (suppl. 2)**:S37–S46.
5.   Schwartlander B, Ghys PD, Pisani E, Kiessling S, Lazzari S, Carael M, *et al*. **HIV surveillance in hard-to-reach populations.** *AIDS* 2001; **15 (suppl. 3)**:S1–S3.
6.   Heckathorn D. **Respondent-driven sampling: a new approach to the study of hidden populations.** *Social Problems* 1997; **44**:174–199.
7.   Heckathorn D. **Respondent driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations.** *Social Problems* 2002; **49**:11–34.
8.   Kish L. *Survey sampling*. New York: Wiley; 1995.
9.   Family Health International. *Sampling techniques for HIV surveillance*. 2005; in press.
10. Semaan S, Lauby J, Liebman J. **Street and network sampling in evaluation studies of HIV risk-reduction interventions.** *AIDS Rev* 2002; **4**:213–223.
11. Thompson SK, Collins LM. **Adaptive sampling in research on risk-related behaviors.** *Drug Alcohol Depend* 2002; **68 (suppl. 1)**: S57–S67.
12. Sutmoller F, de Souza CT, Monteiro JC, Penna T. **The Rio de Janeiro HIV vaccine site-II. Recruitment strategies and socio-demographic data of a HIV negative homosexual and bisexual male cohort in Rio de Janeiro, Brazil.** *Mem Inst Oswaldo Cruz* 1997; **92**:39–46.
13. Villarinho L, Bezerra I, Lacerda R, Latorre Md Mdo R, Paiva V, Stall R, Hearst N. **Vulnerability to HIV and AIDS of short route truck drivers.** *Brazil Rev Saude Publica* 2002; **36 (4 suppl.)**:61–67.
14. Sharma AK, Aggarwal OP, Dubey KK. **Sexual behavior of drug-users: is it different?** *Prev Med* 2002; **34**:512–515.
15. Pal BB, Acharya AS, Satyanarayana K. **Seroprevalence of HIV infection among jail inmates in Orissa.** *Indian J Med Res* 1999; **109**:199–201.
16. Thiede H, Romero M, Bordelon K, Hagan H, Murrill CS. **Using a jail-based survey to monitor HIV and risk behaviors among Seattle area injection drug users.** *J Urban Health* 2001; **78**:264–278.
17. Kassira EN, Bauserman RL, Tomoyasu N, Caldeira E, Swetz A, Solomon L. **HIV and AIDS surveillance among inmates in Maryland prisons.** *J Urban Health* 2001; **78**:256–263.
18. Avila MM, Casanueva E, Piccardo C, Liberatore D, Cammarieri G, Cervellini M, *et al*. **HIV-1 and hepatitis B virus infections in adolescents lodged in 19 security institutes of Buenos Aires.** *Pediatr AIDS HIV Infect* 1996; **7**:346–349.
19. Thaisri H, Lerwitworapong J, Vongsheree S, Sawanpanyalert P, Chadbanchachai C, Rojanawiwat A, *et al*. **HIV infection and risk factors among Bangkok prisoners, Thailand: a prospective cohort study.** *BMC Infect Dis* 2003; **3**:25.
20. Choopanya K, Vanichseni S, Des J, Plangsringarm K, Sonchai W, Carballo M, *et al*. **Risk factors and HIV seropositivity among injecting drug users in Bangkok.** *AIDS* 1991; **5**:1509–1513.
21. Fauziah MN, Anita S, Sha'ri BN, Rosli BI. **HIV-associated risk behaviour among drug users at drug rehabilitation centres.** *Med J Malaysia* 2003; **58**:268–272.
22. Razak MH, Jittiwutikarn J, Suriyanon V, Vongchak T, Srirak N, Beyer C, *et al*. **HIV prevalence and risks among injection and noninjection drug users in northern Thailand: need for comprehensive HIV prevention programs.** *J Acquir Infect Defic Syndr* 2003; **33**:259–266.
23. Department of Control, Minister of Health, China. **National sentinel surveillance of HIV infection in China from 1995 to 1998. Chung-Hua Liu Hsing Ping Hsueh Tsa Chih**. *Chinese J Epidemiol* 2000; **21**:7–9.
24. Ghys PD, Diallo MO, Ettiegne-Traore V, Kale K, Tawil O, Carael M, *et al*. **Increase in condom use and decline in HIV and sexually transmitted diseases among female sex workers in Abidjan, Cote d'Ivoire, 1991–1998.** *AIDS* 2002; **16**:251–258.
25. Risbud A, Mehendale S, Basu S, Kulkarni S, Walimbe A, Arankalle V, *et al*. **Prevalence and incidence of hepatitis B virus infection in STD clinic attendees in Pune, India.** *Sex Trans Infect* 2002; **78**:169–173.
26. Levine WC, Revollo R, Kaune V, Vega J, Tinajeros F, Garnica M, *et al*. **Decline in sexually transmitted disease prevalence in female Bolivian sex workers: impact of an HIV prevention project.** *AIDS* 1998; **12**:1899–1906.
27. Gray JA, Dore GJ, Li Y, Supawitkul S, Effler P, Kaldor JM. **HIV-1 infection among female commercial sex workers in rural Thailand.** *AIDS* 1997; **11**:89–94.
28. Caiaffa WT, Mingoti SA, Proietti FA, Carneiro-Proietti AB, Silva RC, Lopes AC, Doneda D. **Estimation of the number of injecting drug users attending an outreach syringe-exchange program and infection with human immunodeficiency virus (HIV) and hepatitis C virus: the AUDE–Brasil project.** *J Urban Health* 2003; **80**:106–114.
29. Peltzer K, Seoka P, Raphala S. **Characteristics of female sex workers and their HIV/AIDS/STI knowledge, attitudes and behaviour in semi-urban areas in South Africa.** *Curationis* 2004; **27**:4–11.
30. Booth RE, Mikulich-Gilbertson SK, Brewster JR, Salomonsen-Sautel S, Semerik O. **Predictors of self-reported HIV infection among drug injectors in Ukraine.** *J Acquir Immune Defic Syndr* 2004; **35**:82–88.
31. Watters JK, Biernacki P. **Targeted sampling: options for the study of hidden populations.** *Social Problems* 1989; **36**:416–430.
32. MacKellar D, Valleroy L, Karon J, Lemp G, Janssen R. **The young men's survey: methods for estimating HIV sero-prevalence and risk factors among young men who have sex with men.** *Public Health Rep* 1996; **111 (suppl. 1)**:138–144.
33. Muhib FB, Lin LS, Stueve A, Miller RL, Ford WL, Johnson WD, Smith PJ. **A venue-based method for sampling hard-to-reach populations.** *Public Health Rep* 2001; **116 (suppl. 1)**:216–222.
34. Heckathorn D, Semaan S, Broadhead R, Hughes J. **Extensions of respondent-driven sampling: a new approach to the study of infection drug users aged 18–25.** *AIDS Behav* 2002; **6**:55–67.
35. Salganik M, Heckathorn D. **Sampling and Estimation in Hidden Populations Using Respondent Driven Sampling.** *Sociol Methodol* 2004; **34(1)**: 193–240.